

PATENT
OCTEL-00700

**METHOD OF AND APPARATUS FOR IMPROVING PRODUCTIVITY
OF HUMAN REVIEWERS OF AUTOMATICALLY TRANSCRIBED
DOCUMENTS GENERATED BY MEDIA CONVERSION SYSTEMS**

FIELD OF THE INVENTION:

The present invention relates to the field of media conversion systems. More particularly, the present invention relates to an apparatus for and method of improving productivity of human reviewers of transcribed files generated by media conversion systems.

BACKGROUND OF THE INVENTION:

Recent advances in media conversion technology, such as speech-to-text conversion and optical character recognition (OCR), have made it possible for computers to perform tasks such as transcribing messages or documents dictated by a user. Dictation systems for personal computers are now fairly common. Also available are scanning or OCR systems that are capable of converting the content of a printed document into computer readable form. These systems, however, are sometimes inaccurate. The users often need to proof-read and make corrections to the output of these systems.

Also emerging as a new technology is the unified messaging system. The Unified Messenger™, manufactured by Octel Communications Corporation of Milpitas, California, is one such system. These unified messaging systems integrate a user's fax messages, voice messages and e-mails in a single logical mailbox such that they can be accessed by telephone or computer. For instance, the Unified Messenger can convert an e-mail message into a voice message so that it can be read to the user over the telephone. Such a system is described in U.S. Patent 5,557,659, which is incorporated herein by reference.

Data entry personnel are currently utilized for the conversion of an audio message to a text message sent to a user's paging device. In such a system, a caller desiring to send a message, telephones the user's paging service and communicates the message to a data entry employee of the paging service. This employee enters the message into a computer and then transmits it to the user's paging device. The text message entered by the employee of the paging service is

then displayed on the display of the user's paging device. The use of human employees in the text entry and transcription of audio messages is expensive and inefficient. Current automated systems are not reliable and accurate enough to be used for a fully automated messaging or transcription system. Accordingly, what is
5 needed is an automated system which also employs human reviewers to maximize the efficiency and effectiveness of the reviewers.

When documents are dictated and recorded as an audio file, a person transcribing the document plays the audio file and enters text representing the document recorded on the audio tape. The use of human employees in text entry
10 and transcription of audio messages is expensive and inefficient. Automated speech-to-text conversion systems are also currently available to convert audio recordings into a text document. However, such conversion systems are inaccurate, requiring the users to proof-read and make corrections to the output of the systems.

SUMMARY OF THE INVENTION:

An apparatus for improving productivity of human reviewers of transcribed documents generated by media conversion systems includes a server/client network of computers, memories and file systems. The server receives and stores voice
15 files created by users of the system. The server is configured for coupling to a speech-to-text media conversion system to receive converted text files of the audio voice files. The server analyzes the converted text files and routes the converted files to the appropriate reviewers according to an adaptive algorithm. The converted files are displayed on the assigned reviewer's screen at the reviewer's
20 workstation. To aid the reviewer in pinpointing potential errors, the workstation displays different segments of the converted files in different colors to reflect different confidence levels of transcription accuracy. Portions of the original voice message that correspond to the potential errors are played back for the reviewer. The reviewers' workstations also perform productivity enhancing functions such as
25 spelling and grammar checking. After the reviewer has made all the necessary corrections, the reviewed files are transmitted back to the server to be stored and accessed by the users. A user database in the server is also updated to store recurrent user-specific errors corrected by the reviewer. A language analysis
30

system is also disposed to adaptively correct user-specific errors in future reviews according to the information in the user database.

BRIEF DESCRIPTION OF THE DRAWINGS:

Figure 1 illustrates in block diagram form a system of the preferred embodiment of the present invention, including a server, a media conversion system, and a network coupling reviewer workstations, communication devices, and output peripherals to the server.

Figure 2 is a flow diagram illustrating the operations of the preferred embodiment of the present invention when a reviewer logs on.

Figure 3 is a flow diagram illustrating the operations of the server in selecting an appropriate reviewer.

Figure 4 is a flow diagram illustrating the operations of the present invention after a reviewer is selected.

Figure 5 is a flow diagram illustrating the operations of the present invention after the text document is reviewed.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT:

Figure 1 illustrates in block diagram form a system of the preferred embodiment of the present invention. The present invention is directed to an apparatus for and method of improving productivity of human reviewers of automatically transcribed documents generated by media conversion systems. It is the purpose of this patent to cover all media conversion systems including but not limited to speech-to-text, text-to-speech, OCR and hand-writing recognition systems, as well as other information processing systems which may require human supervision. However, for the sake of simplicity, the exemplary system illustrated in Figure 1 and described herein is configured for coupling to only a speech-to-text media conversion system. However, as should be apparent to those skilled in the art, the preferred embodiment of the present invention is easily adapted for use with other types of media conversion systems or information processing systems.

As shown in Figure 1, a speech-to-text media conversion system 125 is coupled to a server 100. The speech-to-text media conversion system 125

preferably includes a conventional computer having a voice processing card such as a SoundBlaster 16 from Creative Labs, and running voice recognition software, such as Naturally Speaking from Dragon Systems, Inc. Alternately, the speech-to-text system 125 is integral to the server 100. The speech-to-text media conversion system 125 automatically converts audio voice files to a text document.

The server 100 is preferably coupled through a telephone switching network 110 to a plurality of reviewer workstations 120 and a plurality of input devices. The input devices include any appropriately configured device through which a user can communicate a message to the server 100. In the exemplary system, illustrated in Figure 1, a telephone set 130, and a dictation machine 131 are included as input devices. Alternately, the input devices are coupled to the server 100 through a PBX or other appropriate network. In a further alternate embodiment, the server 100 is coupled to the reviewer workstation 120 via a computer network such as the Internet. The server 100 is also coupled to a plurality of output peripherals including a fax machine 133 and a printer 132. The reviewer workstations 120 can be located in a centralized location or can be located at the reviewer's office or home and connected to the server 100 through the telephone network by a modem.

Preferably, the server 100 includes an integral language analysis system 101 for improving the accuracy of the text document generated by the speech-to-text media conversion system 125. It is important to note that the language analysis system 101 performs functions different from those of the speech-to-text media conversion system 125. While the speech-to-text media conversion system 125 analyzes and translates the voice files phonetically, the language analysis system 101, such as the Mikrokosmos semantic analyzer from New Mexico State University at Las Cruces, New Mexico, analyzes the context of the resulting text document and corrects phonetically correct but contextually inappropriate words. For example, if the speech-to-text media conversion system 125 converts the sentence: "I red a book", the language analysis system 101 will substitute "read" with "red" in order to correct the resulting text document. In addition, user-specific mistakes that have been corrected in previous reviews of documents submitted by the user are also stored in a user database 102 within the server 100.

Preferably, the language analysis system 101 is disposed to adaptively process the text file according to the information stored in the user database 102.

The user database 102 is also integral to the server 100 and contains information pertaining to the identities of the users. In addition, the user database 102 contains a list of reviewers who have previously reviewed documents submitted by the user. Furthermore, the user database 102 includes a list of preferred reviewers designated by the user. The preferred reviewers are given a higher priority during the assignment process, after the user has submitted a new document.

In the preferred embodiment of the present invention, the server 100 also includes a reviewer database 104. The reviewer database 104 contains a plurality of skill sets for each of the reviewers. A skill set is a list of keywords that a reviewer is familiar with. For instance, for a reviewer proficient in reviewing or transcribing medical documents, his skill set will include common medical terms. The reviewer database 104 is updated when a new reviewer is hired, or when a reviewer acquires a new skill set. The reviewer database 104 also contains other information, such as the number of reviews made and the average time a reviewer spends to finish reviewing a document. Other information useful for evaluating the performance of the reviewer is also stored in the reviewer database 104.

The server 100 also maintains a reviewer log 106 which comprises a list of currently available reviewers. The reviewer log 106 is updated whenever a reviewer logs on or off the server 100, or when a reviewer is assigned a new reviewer package. The server 100 uses this information as a basis for determining whether a reviewer is available to accept a newly submitted document.

The server 100 includes the control processor 108 which is coupled to the language analysis system 101, the user database 102, the reviewer database 104 and the reviewer log 106. The control processor 108 controls the acceptance of a new document from a user and the assignment of the document to a reviewer. The control processor 108 is also coupled to communicate with the network 110 in order to send a document to a reviewer's workstation and to communicate instructions to and receive confirmations from the reviewer.

The server 100 also includes a storage device 103 for storing documents to be sent to or received from reviewers. The storage device 103 is preferably a hard

disk drive. Alternately, the storage device 103 is any appropriately configure media storage device.

After a voice file is created, it is transmitted to the speech-to-text media conversion system 125 for conversion into a corresponding text file, and then to the language analysis system 101. The language analysis system 101 checks the text document for contextually and grammatically incorrect words as described above. The server 100 then determines which one of the currently available reviewers listed in the reviewer log 106 should review the text document. The server 100 makes this decision by first extracting a list of keywords from the text file. The keywords are extracted from the text file by searching the document and comparing it to a keyword database stored in the storage device 103. The server 100 then matches the list of keywords to a skill set within the reviewer database. If the skill set of a reviewer matches the keywords, and if that particular reviewer is available, the server will route the text document file and pointers corresponding to the voice file to the workstation 120 associated with the reviewer. Alternately, the actual voice file is also routed to the reviewer's workstation 120. By assigning text files to reviewers with specialized knowledge and skills in a particular field, the productivity of the reviewers is increased.

To fully maximize the throughput of the reviewers, each of the reviewers' workstations 120 are equipped with a color monitor 122 and a microphone 124. Preferably, the reviewers' workstations 120 also include other conventional input devices such as a keyboard, mouse and a foot pedal. The color monitor 122 is preferred because it is desirable to display the transcribed text file in different colors. In the preferred embodiment, sections of the transcribed text file having a low confidence level are displayed in red, sections of the text having a high confidence level are displayed in green and sections of the text having a medium confidence level are displayed in yellow. Alternately, any desired color combination can be used to highlight the different confidence levels within the text document. With the contrasting colors representing the different confidence levels, the reviewer is able to quickly pinpoint potential errors without carefully reading every single word in the document. Alternately, a monochrome monitor is included at the reviewer's' workstation 120. In this alternate embodiment, sections of the transcribed text file having a low confidence level are highlighted and sections of

text having a high confidence level are displayed normally. In a further alternate embodiment, sections of the text file having a low confidence level are underlined, and sections of the text file having a high confidence level are italicized.

5 The microphone 124 further increases productivity by obviating the need to use a mouse or other conventional input device to navigate around the document. In the preferred embodiment, the workstation 120 is voice activated and controlled by simple voice commands such as "select", "delete", "scroll down", etc. Alternately, the workstation 120 utilizes the keyboard to enter any changes to the text file, and the mouse to select a portion of the text document and the foot pedal to start and stop the replay of the recorded audio voice file.

10 Figure 2 is a flow diagram illustrating the operations of the server in selecting an appropriate reviewer. As shown at the step 200, a user first calls the system through an input device such as the telephone 130 or the dictation machine 131 and records a voice file. The voice file is then transferred to a speech-to-text media
15 conversion system 125 where it is automatically transcribed into a corresponding text file at the step 202. The voice file is then stored in conjunction with the text file in the memory storage device 103. At the step 204, the text file is analyzed by the language analysis system and marked with confidence levels indicating the possibility of errors as described above. The language analysis system 101
20 determines which parts of the text document are likely to be erroneous, and which parts are likely to be correct. The text file is also scanned by the language analysis system 101 at the step 206 to produce a list of keywords corresponding to the text file. At the step 208, the list of keywords corresponding to the text file is compared with the skill sets in the reviewer database 104. The reviewer database
25 104 contains a list of keywords relating to the field with which the reviewer is familiar. By matching the words contained in the skill sets with the words in the list generated by the speech-to-text media conversion system 125, the reviewers whose skill set has a higher correlation with the list of keywords will be accorded a higher priority. Then, the reviewer log 106 is consulted to determine the current
30 availability of the reviewers at the step 210. The available reviewer with the highest priority is selected at the step 212 and a reviewer package containing the voice file and the text file, as well as the criteria for selection is sent to the workstation 120 of the selected reviewer at the step 214. Within the preferred

embodiment of the present invention, the assignment process can be manually overridden such that a user may request a particular reviewer to review his document.

5 In some applications, it is desirable to give a higher priority to a reviewer who has prior experience with the user. In these applications, the system will store a number representative of the number of times a reviewer has reviewed documents from a particular user. This feature can be used in conjunction with the features described above to select the most appropriate reviewer and to maximize the reviewer's review of documents with which he has a familiarity.

10 Figure 3 is a flow diagram illustrating the operations of the preferred embodiment of the present invention when a reviewer logs on to the server 100. As shown, a reviewer logs on to the server 100 through a workstation 120 by entering his name and an identity verifying a password at the step 300. The reviewer is then prompted to download a new reviewer package from the server 15 100 at the step 302. If the reviewer desires to receive a new assignment, the reviewer enters the appropriate commands to download the appropriate files from the server 100 to the reviewer's workstation 120. After a reviewer package is downloaded, the reviewer's workstation 120 displays the criteria used for selecting the reviewer at the step 304. The criteria for selecting the reviewer is displayed so that the reviewer is able to quickly determine if the system has correctly assigned 20 tasks matching his skill set. The reviewer may choose to review the transcription or to reject the assigned task at the step 306. If the task is rejected, the reviewer package is deleted, the server selects a second reviewer and a new reviewer package is transmitted to the workstation 120 corresponding to the second reviewer at the step 308. Once a reviewer accepts the new assignment, the review of the text file is performed at the step 310, as illustrated in the steps included in Figures 4 and 5.

25 Figure 4 illustrates the operations of the present invention after a reviewer is selected. After a reviewer chooses to accept the assignment at the step 400, the text file is displayed in different colors according to the different confidence levels at the step 402, so that the reviewer can quickly pinpoint problematic or low confidence level areas, as described above. The reviewer then selects any appropriate portion of the text within the message using voice activated commands 30

or other conventional input devices at the step 404 during their review. When a section of the text file is selected, preferably a corresponding section of the audio voice file is played for the reviewer at the step 406. The reviewer has the ability through the client computer 120 to control the speed of this audio replay.

5 Alternatively, the reviewer particularly selects portions of the text for which the corresponding audio voice file is played. After listening to the voice file, the reviewer then corrects any inaccuracies at the step 408. In the event that the reviewer cannot determine accurately what was recorded from the voice file because of poor recording quality, the reviewer may establish a connection with the
10 user who submitted the document at the step 410. If the author of the message also has access to a workstation 120, the text file is simultaneously displayed on the author's terminal at the step 412. The reviewer then receives comments from the author and makes further corrections to the file at the step 414.

15 The reviewer then repeats this process and corrects all the inaccuracies within the text file at the step 416. After the reviewer is finished, a check for spelling and grammar mistakes is automatically performed at the step 418, to ensure that the document contains a minimal level of errors.

20 Figure 5 illustrates the operation of the preferred embodiment of the present invention after the review of the text document is complete and all apparent inaccuracies have been corrected. As shown in Figure 5, after the reviewer has finished their review of the document, the edited text document is transmitted to the server 100 to be saved, at the step 500. The edited text document is saved within the media storage device 103. The server 100 also saves information pertaining to the corrections made by the author in the user database at the step
25 502. This information is used in transcriptions of future documents by that author. Then at the step 504, the list of available reviewers is updated to add the reviewer who finished the last review to the list of reviewers available for new assignments. The reviewer database is also updated at the step 506 to store the time necessary to review and correct the last file. Finally, at the step 508, the stored text file is
30 printed out and sent to the author. The text file may also be sent electronically to an intended recipient as an e-mail document.

An apparatus for and method of improving the productivity of human reviewers of computer-processed information has herein been described. This

apparatus and method improves the efficiency of human reviewers and improves accuracy of an automated conversion system. The present invention also enables an integrated messaging system to provide accurate transcriptions of many types of documents including converting voice messages into fax messages or e-mail messages at a low cost. Furthermore, employing the present invention, an integrated messaging system can accurately convert a fax message into a voice message to be retrieved by telephone.

Although the preferred embodiment of the invention uses a speech-to-text system for its implementation, it should be apparent to persons skilled in the art that other information processing functions, such as text-to-speech, OCR or handwriting recognition functions, can also be implemented within the system of the present invention based on the principles herein discussed. Further, there may evolve new message types that would allow similar implementations.

Consequently, this disclosure is not meant to narrow the invention by illustrating the use of the speech-to-text media conversion system as an example. It will be apparent to those skilled in the art that modifications may be made in the embodiment chosen for illustration without departing from the spirit and scope of the invention. Specifically, it will be apparent to one of ordinary skill in the art that the method of the present invention could be implemented in many different ways and the apparatus disclosed above is only illustrative of the preferred embodiment of the present invention.